REVUE DE STATISTIQUE APPLIQUÉE

P. THIONET

Sur certains succédanés du test de Neyman et Pearson

Revue de statistique appliquée, tome 15, n° 2 (1967), p. 19-38

<http://www.numdam.org/item?id=RSA_1967__15_2_19_0>

SUR CERTAINS SUCCEDANES DU TEST DE NEYMAN ET PEARSON

P. THIONET
Faculté des Sciences de Poitiers

PREAMBULE

On se propose de reprendre ici les questions soulevées par la thèse de M. César Lerman (3ème cycle, Paris 1966). Cet excellent travail expose une façon originale de faire un test praticable lorsque celui de Neyman et Pearson est inapplicable en fait (ce qui est fréquent). Nous cherchons à éviter les notations trop générales (et par suite peu élégantes) les raisonnements d'une absolue rigueur (et de ce fait peu commodes à suivre), et d'une façon générale la matière abstraite de la thèse. Nous en conserverons l'exemple numérique (qui a demandé des heures de calcul à la main par centaines) et les concepts essentiels, - en nous excusant de mutiler ainsi l'ensemble de l'œuvre.

Précisons bien tout d'abord que les matériaux dont M. Lerman a besoin se réduisent à une table de logarithmes, une table de l'intégrale de Laplace-Gauss, une table de nombres aléatoires (par exemple de Tippett). Nous y ajouterons nous-mêmes quelques accessoires moins connus.

Pour simplifier, nous nous limiterons au cas des tests d'hypothèses simples. On pourrait se reporter au Chapitre 3 de la Thèse pour voir comment est traité le cas d'une hypothèse simple opposée à une hypothèse composite.

Nous désignerons par n ce que M. Lerman appelle N et vice-versa, ce qui soulagera la mémoire du lecteur, vu que, de cette façon n est (disons) égal à 9 et N égal à 50. On ne trouvera $N < n$ qu'à la fin.

1ère PARTIE :
RAPPEL DU PROBLEME DES TEST D'HYPOTHESES SIMPLES

PROBLEME

Considérons deux lois de probabilités distinctes, dont les fonctions de répartition sont respectivement F(x) et G(x), et un échantillon de n observations $\{x_j\}$. Ces observations sont censées être n valeurs indépendantes prises par une certaine variable aléatoire X, à laquelle on ne sait trop s'il convient d'attribuer la fonction de répartition F(x) plutôt que G(x), ce que nous désignerons pour simplifier par l'hypothèse (F), ou hypothèse zéro, et l'hypothèse (G) ou alternative. A vrai

dire, c'est (F) qu'on a à priori des raisons de tenir pour vraie ; et il faudra que les observations x soient particulièrement nettes pour qu'on accepte (G) au lieu de (F).

Le test sur lequel nous nous fonderons pour choisir (F) ou (G) est celui du "ratio des vraisemblances", et il convient au préalable de définir la vraisemblance.

DEFINITION

On appelle vraisemblance d'une hypothèse A, compte tenu d'un échantillon $\mathcal{S}$, une expression identique (à une constante près éventuellement) à ce qu'on connaît sous le nom de probabilité de l'échantillon $\mathcal{S}$ sous l'hypothèse A (probabilité conditionnelle).

Quand $\mathcal{S}$ n'est pas encore tiré au sort, on parle de probabilité (conditionnelle). Quand $\mathcal{S}$ est donné, la même expression est une "vraisemblance".

Si l'on juge commode dans les calculs de négliger quelques constantes, il ne se peut agir que de constantes numériques (telles que $\sqrt{2\pi}$) qui ne dépendent ni des valeurs x_j de $\mathcal{S}$, ni des paramètres θ de la loi de probabilité : $\theta(A)$.

Explicitons la vraisemblance V pour une loi à densité et pour une loi discrète :

Loi à densité : $f(x) = F'(x)$: on posera $V_F = \prod_j f(x_j)$.

La probabilité de $\mathcal{S}$ sous l'hypothèse F étant $V_F\ dx_1,\ dx_2 \ldots dx_n$.

Loi discrète : La variable X peut seulement prendre l'une des valeurs d'un ensemble discret, que nous supposerons (pour simplifier) être $\{i\} = \{1, 2, 3, \ldots\}$. Posons :

$$P_i = \text{Prob}(X = i)$$

Les observations $\mathcal{S}$ ou $\{x_j\}$ se résument en les nombres de fois que X prend les valeurs 1, 2, 3,... Soit n_i = nombre de fois que X = i, avec :

$$\sum_i n_i = n$$

$$V_F = \prod_i P_i^{n_i}$$

RATIO DES VRAISEMBLANCES

Considérons alors les "vraisemblances" V_F et V_G des hypothèses (F) et (G) pour le même échantillon $\mathcal{S}$.

Si l'on avait $V_F > V_G$ - si (F) était "plus vraisemblable" que (G), il n'y aurait pas de problème. C'est au contraire parce que (compte-tenu de nos observations $\mathcal{S}$) on a $V_F \leq V_G$ que (F) "semble moins vraisemblable" que (G), ce qui est en contradiction avec ce que nous pensions être correct.

Jusqu'où peut-on aller trop loin ?

Jusqu'où peut-on tolérer que $V_F/V_G = U$ soit plus petit que 1 ?

Il est fort naturel qu'on décide de renoncer à (F), au profit de (G), si (décidément) nous trouvons V_F/V_G trop voisin de O.

PASSAGE AUX LOGARITHMES (népériens)

Posons $\Lambda = \text{Log } V$, il est clair que

$$U = \frac{V_F}{V_G} < U_o \quad \Longleftrightarrow \quad \text{Log } V_F - \text{Log } V_G < \text{Log } U_o \quad \Longleftrightarrow \quad \Lambda_F - \Lambda_G \leqslant \lambda_o$$

On pose souvent $\lambda = \Lambda_F - \Lambda_G$; ainsi, on ne s'intéresse qu'au cas où λ est négatif ($U < 1$). La statistique $(-\lambda)$ est positive ; mais on peut dire que :

{ Tant que $-\lambda$ n'est pas trop grand, on préfère accepter (F)
{ Si pourtant, $-\lambda$ est trop grand, on rejette (F) pour accepter (G).

Remarque 1 : En pratique, il est plus commode de "tester" l'écart entre les logarithmes des vraisemblances que le "ratio" des vraisemblances.

Remarque 2 : Sous certaines conditions (sur lesquelles nous reviendrons) -2λ ou -4λ a une distribution de probabilités peu différente de celle du χ^2 de Pearson, c'est pourquoi nous avons tant insisté sur le fait que $-\lambda$ est positif.

Expression du test : a) 2 Lois à densité f(x), g(x)

$$U = \frac{V_F}{V_G} = \prod_j \frac{f(x_j)}{g(x_j)}$$

$$\lambda = \Lambda_F - \Lambda_G = \sum_j L \frac{f(x_j)}{g(x_j)} = \sum_j Lf(x_j) - Lg(x_j)$$

b) 2 Lois discrètes : (p_i), (q_i)

$$U = \frac{V_F}{V_G} = \prod_i \left(\frac{P_i}{q_i}\right)^{n_i}$$

$$\lambda = \Lambda_F - \Lambda_G = \sum_i n_i L\left(\frac{p_i}{q_i}\right) = \sum_i n_i Lp_i - n_i Lq_i$$

USAGE DE LA VARIABLE λ

Le choix de F(x) et G(x) étant fait, avant qu'on se propose de choisir entre F et G, l'expression de λ (ou de U) est connue ; connaissant les observations $\mathcal{X}$, λ est un nombre. C'est la valeur prise par une certaine variable aléatoire que nous désignerons encore par λ.

Si (F) est correcte, λ a pour fonction de distribution $H(\lambda/F) = H_F(\lambda)$.

Si (G) est correcte, λ a au contraire la fonction de distribution $H(\lambda/G) = H_G(\lambda)$.

Nous n'aurons pas à nous occuper de ce qui se passe si (F) et (G) sont fausses l'une et l'autre.

DEFINITION DES DEUX RISQUES

On pose d'habitude :

$\alpha = \text{Prob}(\lambda < \lambda_0 | F)$ en admettant que, si λ est inférieur à un certain λ_0, on renonce à la loi (F) au profit de (G).

$1 - \beta = \text{Prob}(\lambda < \lambda_0 | G)$ ... (puissance du test λ).

Ainsi α est la probabilité qu'on rejette (F) alors que (F) est vraie ;

β est la probabilité qu'on accepte (F) alors que (G) est vraie ;

α et β sont dits les risques de 1ère et 2ème espèces que comporte la règle de décision :

$$\lambda < \lambda_0 \Longrightarrow \text{rejeter F et accepter G}$$

Remarque : Plus couramment, on se donne α ; d'où λ , d'où β. En effet : par définition :

$$\alpha = H(\lambda_0 | F) = H_F(\lambda_0) \Longrightarrow \lambda_0 = H_F^{-1}(\alpha)$$

$$1 - \beta = H_G(\lambda_0) \Longrightarrow \beta = 1 - H_G[H_F^{-1}(\alpha)]$$

(le symbole H_F^{-1} désignant la fonction inverse de H_F, dont on suppose l'existence).

Optimalité du test λ (Théorème de Neyman et Pearson). Soit T une statistique quelconque. Soit $H_F^*(t) = \alpha$ donné ; supposons qu'on veuille réduire β le plus possible : en désignant par H_F^* la fonction de distribution de T (la probabilité de $t < T$) sous l'hypothèse F, et par β l'expression $1 - H_G^*(t)$.

On démontre que (sous certaines conditions à préciser) c'est la statistique λ qui, parmi toutes les statistiques T, (α, F, G et $\{x_j\}$ restant les mêmes bien entendu) réalise le plus petit β, c'est-à-dire le moindre risque de 2ème espèce.

En conséquence, il est des tests "aussi bons" que λ dans telle ou telle hypothèse ; il ne peut y en avoir de meilleur. Le fait est pourtant qu'on ne l'utilise guère.

DIFFICULTES D'APPLICATION DU TEST λ

La difficulté n'est pas de calculer le nombre λ : C'est de savoir s'il est vraiment trop petit, c'est-à-dire de le placer sur une échelle de valeurs. Autrement dit, c'est la fonction $H_F(\lambda)$ qui n'est généralement pas accessible numériquement.

Si l'on désire en outre connaître le risque de 2ème espèce β, on aura besoin de la fonction $H_G(\lambda)$. Si les fonctions F et G sont prises dans une même famille de fonctions différant par leur(s) paramètre(s), il peut se faire que le calcul de H_G se déduise simplement de celui de H_F ; mais H_F vient de nous arrêter. Examinons en détail un cas particulier,

celui des lois de Darmois-Koopman-Pitman (lois DKP), famille étendue qui comprend en fait presque toutes les lois de probabilités utilisées.

CAS DES LOIS D.K.P.

Soit : $\text{Log } f(x) = \alpha(\theta) . a(x) + \gamma(\theta) + A(x)$

$\text{Log } g(x) = \beta(\omega) . b(x) + \delta(\omega) + B(x)$

où, θ, ω sont des paramètres connus, donc α, β, γ, δ, des nombres.

On a donc :

$$\lambda = \alpha \sum_j a(x_j) + \beta \sum_j b(x_j) + \sum_j A(x_j) - \sum_j B(x_j) + n(\gamma - \delta)$$

Si a(x), b(x), A(x), B(x) sont des fonctions assez simples, il ne doit pas être impossible de trouver la loi de répartition sous F, puis sous G, des 4 convolutions :

$$\Sigma a(x_j) \qquad \Sigma b(x_j) \qquad \Sigma A(x_j) \qquad \Sigma B(x_j)$$

mais comme il s'agit de 4 variables aléatoires non indépendantes, on n'en peut déduire directement la loi de répartition de λ.

CAS DE DEUX LOIS D'UNE MEME FAMILLE A UN PARAMETRE

Pour simplifier, conservons l'expression de Log f(x), mais supposons g(x) donné comme suit :

$$\text{Log } g(x) = \alpha(\theta') \; a(x) + \gamma(\theta') + A(x)$$

d'où :

$$\text{Log } f(x) - \text{Log } g(x) = [\alpha(\theta) - \alpha(\theta')] \; a(x) + [\gamma(\theta) - \gamma(\theta')]$$

Si donc nous savons calculer la fonction caractéristique de a(x), soit :

$$\varphi_a(t)$$

alors celle de λ dépendra seulement de $\varphi_a^n(t)$. Alors on songera :

- soit à obtenir H par l'inversion de Fourier (possibilités pratiques médiocres) ;

- soit à reconnaître la fonction caractéristique d'une loi connue (rares succès) ;

- soit à approcher la loi de $\Sigma a(x_j)$ par une loi de Laplace-Gauss en vertu du théorème central limite (ce qui suppose n grand).

Conclusion : Il reste de beaux rôles à jouer pour des succédanés du test de Neyman et Pearson λ, si possible à peine moins puissants mais se prêtant au calcul effectif jusqu'à l'obtention du résultat : garder ou ne pas garder (F).

2ème PARTIE

APPLICATION A UN EXEMPLE

Nous reprendrons tout simplement l'exemple pris par Lerman (Ch. 4) et nous essaierons de lui appliquer le test de Neyman et Pearson classique. Il se trouve que F et G sont deux lois D.K.P., sans être de la même famille (au sens restreint donné ci-dessus au mot famille). Nous avons pu mener très loin les calculs.

Exemple :

$F(x) = 1 - e^{-x}$	$f(x) = e^{-x}$	$Lf(x) = -x$
$G(x) = 2\Phi(\sqrt{x}) - 1$	$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}}$	$Lg(x) = -\frac{x}{2} - \frac{1}{2} Lx + C$

avec :

$$\Phi(y) = \int_{\infty}^{y} e^{-\frac{t^2}{2}} dt/\sqrt{2\pi}$$

d'où :

$$-\lambda = \sum_j \frac{1}{2} (x_j - Lx_j) - \frac{n}{2} L2\pi$$

Remarquons en passant que $E X = 1$ dans (F) comme dans (G) ; car la loi (G) est une loi gamma pour la variable $Z = X/2$, et par suite $EZ = h = \frac{1}{2}$

$$E(Z) = \frac{\Gamma(h+1)}{\Gamma(h)} \qquad \text{avec ici} \quad h = \frac{1}{2} .$$

En fait les données $\mathcal{S}$ dont on s'occupera auront pour moyenne 1,04 ce qui explique le choix de F et G ayant l'une et l'autre l'espérance mathématique 1.

CALCUL DES FONCTIONS GENERATRICES DE MOMENTS DE $-\lambda$

On aura :

$$E(\exp(-\lambda t)) = \exp\left[-\frac{n}{2} (L2\pi) t\right] . \left\{E\left[\exp \frac{}{2} (x - Lx)\right]\right\}^n$$

Commençons par calculer :

$$\varphi(t) = E \exp[t(Lx - x)] = E(x^t e^{-tx})$$

Hypothèse F :

$$\varphi(t) = \int_0^\infty e^{-(1+t)x} x^{(1+t)-1}\, dx$$

$$= \frac{\Gamma(1+t)}{(1+t)^{1+t}}$$

Hypothèse G :

$$\varphi(t) = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}+t\right)x} x^{-\left(\frac{1}{2}+t\right)-1}\, dx$$

$$= \frac{\Gamma\left(\frac{1}{2}+t\right)}{\sqrt{2\pi}\left(\frac{1}{2}+t\right)^{\frac{1}{2}+t}}$$

Il sera plus commode de poser $L\varphi(t) = \psi(t)$, fonction génératrice des cumulants. On a renoncé à reconnaître une variable aléatoire connue ou à inverser $\psi(t)$. En revanche le calcul des moments de $-\lambda$ est apparu praticable ; il suppose qu'on sache dériver la fonction Gamma, les résultats nécessaires sont justifiés en annexe.

On obtient ainsi,

sous l'hypothèse F :

$$\left|\begin{array}{l} E(Lx - x) = -(1+\gamma) \qquad (\gamma = 0{,}5772). \\ V(Lx - x) = \dfrac{\pi^2}{6} - 1 = 0{,}645 \end{array}\right.$$

sous l'hypothèse G :

$$\left|\begin{array}{l} E(Lx - x) = -(1 + \gamma + L\,2) = -(1+\gamma) - 0{,}69 \\ V(Lx - x) = \dfrac{\pi^2}{2} - 2 = 2{,}935 \end{array}\right.$$

Le fait est que l'écart-type de λ sera (sous G) bien supérieur (le double) de celui obtenu sous (F), outre la translation infligée à l'espérance mathématique.

Comme la grande ressource est, en définitive, d'assimiler ces distributions de λ à deux lois de Laplace-Gauss, il est nécessaire de se rendre compte de la valeur prise par les moments d'ordre 3 et 4 de $(Lx - x)$. Poussant plus loin les dérivations de $\psi(t)$, on obtient les cumulants suivants :

(Hypothèse F) $\left|\begin{array}{l} K_3 = -1{,}40 \\ K_4 = +4{,}41 \end{array}\right.$ au lieu de 0 pour une loi de Laplace-Gauss

ce qui n'est guère encourageant (surtout si n est de l'ordre de 10).

Application numérique : $n = 9 \quad \log 2\pi = 0,30103 + 0,49715 = 0,79812$

x_j	$\log x_j$
0,052	$\bar{2},71600$
0,068	$\bar{2},83251$
0,165	$\bar{1},21748$
0,195	$\bar{1},29003$
0,459	$\bar{1},66181$
0,772	$\bar{1},88762$
1,71	0,23300
1,98	0,29667
4,00	0,60206
9,401	$\bar{4},73718$
	$-3,26292$

$\text{Log } X = \log X / 0,43429 \left| \begin{matrix} -3,26892/0,43429 = -7,51 \\ 0,79818/0,43429 = 1,84 \end{matrix} \right.$

$$\lambda = \frac{n}{2} L\, 2\pi + \frac{1}{2} \Sigma(Lx - x) = \underline{\overline{-0,17}}$$

Pour avoir une échelle de grandeur, comparons λ à :

$$E\lambda = \frac{n}{2} L\, 2\pi + \frac{n}{2} E(Lx - x)$$

Autrement dit, comparons :

$$-\Sigma(Lx - x) = -7,51 - 9,40 = -16,91$$

et $\qquad \Delta = -2,7$

$$nE(Lx - x) = 9(-1,5772) = -14,195$$

L'écart entre ces deux grandeurs est faible, si l'on songe que la variance du premier est :

$$V(\Sigma Lx - x) = nV(Lx - x) = 9\left(\frac{\pi^2}{6} - 1\right) = 5,81 = (2,41)^2$$

Quelle que soit la vraie distribution $H_F(\lambda)$, il est clair qu'un tel écart Δ, de l'ordre de un écart-type n'est pas significatif, c'est-à-dire que : la probabilité d'avoir $-\lambda$ supérieur à 0,17 est supérieure à 5 %, à 10 %, à 20 % même (du moins, c'est là une opinion de statisticien). C'est bien d'ailleurs ce que trouve de son côté Lerman, par un calcul tout différent.

Quant à la distribution de λ sous l'hypothèse G, elle est encore plus décevante :

$$nE(Lx - x) = -14,195 - 9.L2 = -20,4 \Longrightarrow \Delta' = 3,5$$

avec un écart-type dépassant 5.

En résumé : avec le test "le plus puissant" (celui de Neyman et Pearson), il apparaît que les lois (F) et (G) sont trop voisines et l'échantillon $\mathcal{S}$ est trop petit pour qu'on puisse faire un choix entre F et G. Un calcul grossier (voir annexe) tend à placer vers n = 200 l'effectif convenable pour $\mathcal{S}$.

3ème PARTIE

LA METHODE DE LERMAN

La méthode qui va être à présent décrite fait une synthèse d'idées très naturelles.

1 - PASSAGE DE LOIS F, G CONTINUES A DES LOIS DISCRETES qui en soient une bonne approximation. L'idée de substituer à une loi F à densité une loi discrète (c'est-à-dire un ensemble de N urnes) est celle qui préside au test de χ^2 de la qualité de l'ajustement ; on sait, en effet, que les tests de "goodness of fit" qui laissent à F son caractère continu sont d'un emploi pratique assez délicat, exception faite du test de Fisher-Neyman.

$$\pi = -2 \sum_j \text{Log } F(x_j) \qquad \text{ou} \qquad \pi' = -2 \sum_j \text{Log } [1 - F(x_j)]$$

Le test destiné à comparer (F) et (G) va donc comporter un découpage de l'intervalle de variation de X en N intervalles partiels (i = 1, 2, . . . , N) ; on posera :

$$\Delta F_i = P_i \ , \qquad \Delta G_i = q_i \ , \qquad \Sigma P_i = \Sigma q_i = 1$$

ΔF_i étant l'accroissement subi par F sur l'intervalle (i) (idem pour G). A la fonction F ou G est substituée la fonction en escalier $F^{(N)}$ ou $G^{(N)}$

$$F^{(N)}(x) = P_1 + P_2 + \ldots + P_{i-1} \qquad \text{si} \qquad x \in (i)$$

$$G^{(N)}(x) = q_1 + q_2 + \ldots + q_{i-1}$$

Si N est grand, (pratiquement, on a utilisé N = 50) les fonctions F et $F^{(N)}$ ne diffèreront guère ; - de même pour G et $G^{(N)}$.

Il a été établi (cf. Thèse Lerman) que le découpage tel que $\Delta F_i = 1/(N+1)$ réalisait un certain optimum. Bien entendu le même découpage fournit des ΔG_i inégaux entre eux et l'optimum, s'il concerne la distribution de λ sous l'hypothèse F ne peut s'étendre à celle du même λ sous G.

2 - SIMULATIONS D'ECHANTILLONNAGE (AU MOYEN DE NOMBRES AU HASARD

La méthode dite de Monte-Carlo comporte le plus souvent l'usage d'un ordinateur. Dans le cas présent, on s'est contenté de tirer M échantillons de n nombres avec une table de nombres aléatoires (en fait n = 9, M = 250). Il s'agissait de nombres à 2 chiffres, de 00 à 99 ; ceux compris de 50 à 99 étaient convertis en nombres allant de 00 à 49.

Ces nombres désignaient les intervalles (i) prélevés parmi les N définis ci-dessus ; soit M échantillons indépendants de n intervalles, avec les n valeurs ΔG_i distinctes correspondant à ces intervalles. Alors la statistique $\lambda = \text{Log } (V_F - V_G)$ devient :

$$\lambda = \Sigma n_i \ \text{Lp}_i - n_i \ \text{Lq}_i = \Sigma n_i \ \text{Log } \frac{1}{N} - \Sigma n_i \ \text{Log } \Delta G_i$$

Autrement dit :

$$-\lambda = \Sigma n_i \ \text{Log } \Delta G_i + n \ \text{Log } N \qquad (n = \Sigma n_i)$$

On calcule numériquement les M valeurs de λ (une pour chaque échantillon), on les classe, on en trace le graphique des fréquences cumulées $H_F^M(\lambda)$:

Cette distribution empirique donne une bonne idée de $H_F(\lambda)$ si le nombre M d'échantillons est suffisamment grand, en vertu du théorème bien connu de Glivenko-Cantelli ou Théorème Central Statistique.

Dans l'exemple traité (cf. 2ème Partie ci-dessus), on a utilisé la Table de logarithmes pour calculer les limites F des intervalles (i) de variation de x, c'est-à-dire pour résoudre les équations :

$$F(\xi_1) = 0,02 \; ; \; F(\xi_2) = 0,04 \; ; \; F(\xi_3) = 0,06 \; \ldots$$

Puis on a obtenu $G(\xi_1)$, $G(\xi_2)$, $G(\xi_3)$,... à l'aide de la table G, c'est-à-dire de la table Φ de l'intégrale de Gauss ; d'où :

$$\Delta G_1 = G(\xi_1) \; ; \; \Delta G_2 = G(\xi_2) - G(\xi_1) \; ; \text{ etc.}$$

Puis $\log \Delta G_1$, $\log \Delta G_2$,... (décimaux) ; enfin on calcule :

$$-\lambda = \frac{(\Sigma n_i \log \Delta G_i + n \log N)}{0,43429}$$

Pour simplifier les calculs, on a arrondi les ΔG_i le plus possible et conservé le plus petit nombre possible de $\log \Delta G_i$ distincts (soit onze sur cinquante) ; l'opérateur connaît par cœur les onze valeurs et travaille sans consulter la table des $\log \Delta G_i$.

Mais tous ces moyens sont employés au service d'une idée nouvelle, intéressante mais qui n'était pas (à notre avis) indispensable : Celle d'un espace échantillon tronqué. C'est là l'élément théorique le plus original de la méthode.

3 - ESPACE ECHANTILLON TRONQUE

Supposons qu'on tire au sort n = 2 intervalles (i) avec probabilités égales 1/N. L'espace échantillon est constitué par N^2 points d'abscisses (1,2...N) et d'ordonnées (1,2,...N). L'échantillon est l'un de ces N^2 points (au hasard). La troncature consiste à supprimer les N points situés sur la diagonale principale.

Plus généralement, le tirage des n intervalles a lieu sans remise (est exhaustif). La troncature n'est pas un détail négligeable : sur N^n points de l'espace-échantillon, on n'en conserve que $N(N-1)/(N-2)\ldots(N-n+1)$ (chaque point conservé figure alors n! fois pour une, par raison de symétrie).

Par exemple pour N = 50, n = 9, on a :

$$f = \frac{N(N-1)\ldots(N-n+1)}{N^n} = 0,4655 \; :$$

la troncature supprime donc plus de 53 pour cent des points de l'espace échantillon.

Initialement, les n valeurs échantillons $\{x_j\}$ étant données, on a choisi N assez grand pour qu'aucun des N intervalles découpés ne renferme plus d'un x_j. On a ainsi un échantillon de n intervalles, ou de n nombres i-1 = 00,..., 49 (en retranchant 1 au numéro de l'intervalle).

Cet échantillon est en un sens comparable aux M qu'on a tirés au sort ci-dessus. Seulement les difficultés théoriques apparaissent à la réflexion.

Difficultés :

a) Il est bien connu que, lorsqu'on applique le test de χ^2, il ne faut pas "tricher" et choisir un découpage de l'intervalle où varie x tel que χ^2 s'en trouve "amélioré". De même ici, le découpage en ΔF_i et ΔG_i reste arbitraire (sinon qu'il est "optimal" pour $\Delta F_i = 1/N$).

Et même, le choix arbitraire de N (avec $\Delta F_i = 1/N$) peut infléchir λ plus ou moins. Est-il tellement légitime de tronquer ainsi l'espace-échantillon ? Il semble que ce soit là plutôt un souci extrême de rigueur dans l'application de certaines règles (qu'en fait on ne respecte guère) concernant l'indépendance du choix des méthodes statistiques vis à vis des observations à traiter.

b) Les échantillons tirés "sans remise" introduisent une corrélation négative entre les n valeurs log ΔG_i tirées ; on a bien toujours :

$$E(-\lambda) = E(\Sigma n_i \text{ Log } \Delta G_i) + n \text{ Log } N$$
$$= nE(\text{Log } \Delta G) + n \text{ Log } N$$

mais la variance est modifiée et devient :

$$Y(-\lambda) = n\, V(\text{Log } \Delta G) \frac{N - n}{N - 1}$$

Ainsi la variance est réduite quelque peu (avec N = 50, n = 9, $(N - n)/(N - 1)$ est 41/49, soit 16 % de réduction). Il s'en suit que le test λ est plus efficace de ce fait, dans le rapport $(N - 1)/(N - n)$ (gain d'efficacité de 20 %) ; c'est excellent.

c) En contrepartie, la distribution de λ sous l'hypothèse (G) donne quelques inquiétudes. Il n'est pas question de l'obtenir par simulation, sinon en repartant de zéro (avec $\Delta G_h = 1/N' = q_h$) : on pourrait tirer des échantillons d'intervalles i avec des probabilités ΔG_i inégales, mais avec remise (les tirages sans remise en pareil cas sont encore une énigme de la théorie des sondages).

d) Pour en revenir à $\Sigma n_i \text{ Log } \Delta G_i$, qu'on peut écrire $\Sigma \text{Log } \Delta G_j$ (puisque $n_i = 1$ si $x_j \in (i)$ et $n_i = 0$ autrement), il est clair que cette expression suit une loi asymptotique de Laplace-Gauss si n est grand ; mais le théorème central limite ne s'appliquant plus (non indépendance des ΔG_j), il a fallu établir le fait directement, grâce au Théorème de Wald et Wolfowitz, assez utile en pareil cas (voir Fraser, Non parametric methods in statistics, 1957 - par exemple).

Ainsi, les parties les plus difficiles et les plus mathématiques de la thèse n'auraient plus de portée pratique si les valeurs de n_i pouvaient devenir égales à 2, 3, etc.

e) Or, il est certain que le gain de variance (du au tirage sans remise des échantillons) ne peut être bien grand ; car cela supposerait n voisin de N, ce qui rend improbable l'existence de l'échantillon initial $\mathcal{S}$ tel que $n_i = 0$ ou 1 (à moins, bien sur, qu'on ne triche en découpant astucieusement des intervalles inégaux ΔF_i ; ceci entraînerait des erreurs, comme il est dit au § c) ci-dessus).

En somme, si n est petit à côté de N, on gagne fort en simplicité (et on perd peu en précision) à faire M tirages indépendants de n tirages indépendants d'intervalles équiprobables ; - on conserve de la méthode de Lerman :

- la substitution de lois discrètes (p_i, q_i) aux lois continues (F, G),

- la simulation en vue d'approcher la fonction $H_F(\lambda)$.

f) La puissance du test λ a été évaluée sans chercher à simuler $H_G(\lambda)$ mais en usant d'une autre définition que celle donnée dans la 1ère Partie (où la puissance est $1 - \beta = H_G[H_F^{-1}(\alpha)]$).

Alors que la proportion d'échantillons tirés (parmi les M tirés) pour lesquels U_o dépasse U estime la valeur de α, la somme des V_G correspondantes rapportée à la masse totale des V_G estime $1 - \beta$; or, $V_G = V_F/U$ est proportionnelle à $1/U$ quand V_F est constant.

Par exemple, sur M = 250 cas, il en est 237 pour lesquels $1/U$ atteint $151,5.10^{-17}$ et 238 pour lesquels $1/U$ atteint $171,4.10^{-17}$. On choisit $\alpha = 5\ \%$ et on a :

$$250 \times (1 - 0,05) = 237,5$$

Les valeurs 151,5 et 171,4 donnent en moyenne (et au facteur 10^{-17} près) une idée de ce que peut être le seuil $1/U = e^{-\lambda}$ significatif au niveau 5 %.

On peut trouver beaucoup moins naturel d'estimer la valeur correspondante de $1 - \beta$ comme suit : la somme des 250 valeurs calculées de $e^{-\lambda}$ est $10.575.236\,(10^{-17})$.

La somme des 13 plus grandes de ces valeurs est $4.287.300\ (10^{-17})$.

La puissance du test (pour $\lambda = 5\ \%$) est :

$$1 - \beta = \frac{4.287.200}{10.575.336} = 40,5\ \%$$

Expliquons cette technique :

En fait, si l'on définit une région critique Ω par l'inéquation $v < v_o$, dans l'espace où est le point échantillon $\{x_j\}$ ou M, on a ainsi (dans l'espace) :

$$\alpha = \int_{\Omega} V_F\, dM \quad , \quad 1 - \beta = \int_{\Omega} V_G\, dM \quad \text{avec} \quad \begin{cases} \int_{\Omega + \bar{\Omega}} V_F\, dM = 1 \\ \int_{\Omega + \bar{\Omega}} V_G\, dM = 1 \end{cases}$$

Disposant de M valeurs équiprobables de V_F et de V_G, on estime $(1 - \beta)$ par le ratio :

$$\frac{\Sigma_{\Omega}\, V_G}{\Sigma_{\Omega + \bar{\Omega}}\, V_G}$$

et bien entendu toutes les V_F sont égales (à N^{-n}) ; d'où :

$$\frac{\Sigma_{\Omega}\, V_F}{\Sigma_{\Omega + \bar{\Omega}}\, V_E} = \frac{\alpha M}{M} = \alpha$$

en appelant αM le nombre des V_F telles qu'on ait $\frac{V_F}{V_G} = v < v_o$, c'est-à-dire

$$V_G > \frac{V_F}{v_o} = \frac{1}{N^n v_o}$$

Inversement, se donnant α (disons $\alpha = 5\ \%$), donc αM, cette inéquation définit v_o (d'où $\lambda_o = \text{Log } v_o$). Ainsi est justifiée la technique d'évaluation de $(1 - \beta)$ de Lerman.

g) Nous ne voyons absolument aucun motif de ne pas appliquer les considérations du § f à l'espace échantillon dans son intégralité au lieu de les restreindre à l'espace tronqué. L'évaluation de $(1 - \beta)$ est valable du moment que les M échantillons sont équiprobables et les V_F égales entre elles.

4ème PARTIE

AUTRES SUCCEDANES DU TEST DE NEYMAN ET PEARSON

1 - RECOURS A M ECHANTILLONS NON INDEPENDANTS

a) Nous ferions en revanche une sérieuse économie si nous pouvions utiliser systématiquement tous les échantillons possibles correspondant à un minimum de données. Par exemple si nous ajoutons une $n + 1^{\text{è}}$ valeur de x, nous obtenons $n + 1$ échantillons possibles ; plus généralement avec k données x supplémentaires, on peut composer C^n_{n+k} échantillons distincts, de n éléments, tous équiprobables. On arriverait vite à 250 et plus :

$$C^9_{12} = 220 \ ; \ C^9_{13} = 715$$

Il suffirait donc de 4 valeurs supplémentaires de x pour obtenir 714 nouveaux échantillons. Il serait assez facile d'obtenir les 714 valeurs de λ, etc. etc., d'où $v_o \lambda_o \beta$. Mais il est assez clair que jamais (ou presque) le λ initial ne diffèrerait significativement des autres, il faut trouver autre chose.

b) Prenons donc 12 (ou 13) nouvelles valeurs de x et formons avec elles les 220 (ou 715) nouveaux échantillons. On peut cette fois leur comparer le λ initial, évaluer v_o λ_o et β.

Je pense que le procédé est aussi correct que celui consistant à tirer M échantillons indépendants. Il est beaucoup plus économique. Il ne peut certainement pas être aussi précis : car les erreurs d'échantillonnage ne se compensent pas quand on tire 12 ou 13 valeurs, comme si l'on en tirait $250 \times 9 = 2.250$.

Il est remarquable que cette façon de faire n'ait pas été envisagée, alors que les "Monte-Carlo" automatiques utilisent de tels échantillons.

c) Or, il s'agit au fond de comparer l'échantillon initial $\mathcal{E}$ à un échantillon $\mathcal{E}'$ que l'on sait (cette fois) tiré dans (F) : un problème des 2 échantillons se pose alors : peut-on accepter que $\mathcal{E}$ et $\mathcal{E}'$ proviennent tous deux de (F) ? ou (alternative) que $\mathcal{E}$ provienne de (G) ?

Nous n'aborderons pas ce problème, qui mérite une étude spéciale.

2 - RECOURS A UN χ^2 (n RESTANT PETIT)

Nous avons pensé que la distribution de $-\lambda$ pouvait être exprimée à l'aide d'un χ^2 avec une approximation suffisante tant que le théorème central limite ne peut jouer. Nous étions incité à cette tentative par le livre de Kullback (Information Theory and Statistics).

Reprenant l'exemple de la 2ème partie (voir annexe § 6), nous avons constaté que -4λ ressemblait à un χ^2 possédant 1,3 n degrés de liberté à condition d'en corriger convenablement l'espérance mathématique (la ressemblance concerne les cumulants d'ordre 2, 3 et 4, donc aussi les moments). Il s'agit de la distribution de -4λ sous l'hypothèse (F) ; rien n'empêcherait de tenter le calcul dans l'hypothèse (G).

En revanche on ne saurait transposer la méthode à des lois (F) et (G) un peu générales.

3 - RECOURS A UN χ^2 (n DEVENANT GRAND)

3.1. - Si n est un peu grand, la méthode LERMAN est totalement impraticable puisqu'elle suppose N beaucoup plus grand encore que n.

En revanche, chacun aura l'idée de substituer à F et G deux distributions discrètes $p_i = \Delta F_i$, $q_i = \Delta G_i$ résultant du découpage en N segments de l'intervalle de variation de x. Le test de Neyman et Pearson appliqué à ces distributions est :

$$\lambda = \Sigma n_i \, L\left(\frac{p_i}{q_i}\right)$$

où (n_i) a une distribution multinomiale : à savoir (p_i) sous (F), ou bien (q_i) sous (G).

La distribution de λ est la projection de celle de (n_i) sur la direction fixe de paramètres directeurs $L(p_i/q_i)$.

Nous avions espéré que -2λ serait un χ^2 asymptotiquement, en nous référant à Kullback (Information theory and Statistics) qui fait allusion à une propriété de cette nature établie par Wilks (1938). Ceci concerne en fait le cas où l'une des lois a ses paramètres estimés sur l'échantillon, les deux lois appartenant à la même famille. Ici au contraire, les 2 lois ont leurs paramètres estimés sur l'échantillon mais dans deux familles différentes (et d'ailleurs on ne tient pas compte en fait de ces estimations).

Nous montrerons cependant (en Annexe) que, si n est devenu plus grand que N (au point qu'aucun n_i ne soit nul), alors 2λ est la différence entre deux χ^2 à (N - 1) degrés de liberté. D'ailleurs ceci n'est actuellement utilisable que dans ces cas très particuliers.

3.2. - Enfin si n devient vraiment grand, le théorème central limite s'applique :

$$\lambda = \prod_n a_j$$

est la somme de n valeurs de la variable $a = L(p_i/q_i)$ tirées (avec remise) dans une urne, avec des probabilités P_i , (F) ou bien q_i , (G).

Donc λ/n possède à la limite une distribution de Laplace-Gauss de moyenne μ_1 et de variance σ^2/n, avec :

$$\mu_1 = \Sigma p_i \, L(p_i/q_i), \text{ sous (F) ; } \Sigma q_i \, L(p_i/q_i), \text{ sous (G)}$$

$$\sigma^2 = \mu_2 - \mu_1^2$$

$$\mu_2 = \Sigma p_i \, [L(p_i/q_i)]^2, \text{ sous (F) ; } \Sigma q_i \, [L(p_i/q_i)]^2, \text{ sous (G)}$$

ANNEXE

1 - On trouve dans Whittaker et Watson, Modern Analysis (1952) page 236 (Ex. 1) $\Gamma'(1) = -\gamma$ avec $\gamma = 0{,}5772$ (constante d'Euler)

page 241 - § 12.16 $$\frac{d}{dz} \operatorname{Log} \Gamma(1+z) = -\gamma + \frac{z}{1(z+1)} \quad \frac{z}{2(3+2)} + \ldots$$

$$\frac{d^2}{dz^2} \operatorname{Log} \Gamma(1+z) = \frac{1}{(z+1)^2} + \frac{1}{(z+2)^2} + \ldots$$

Pour $z = 1$, on retrouve $\Gamma'(1) = -\gamma$ [$\Gamma(1) = 1$ bien entendu]

$$[\operatorname{Log} \Gamma(1)]' = -\gamma$$

$$[\operatorname{Log} \Gamma(1)]'' = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots$$

Par ailleurs, on trouve (page 163, ex. 2) la série trigonométrique

(T) $$\frac{\pi^2}{12} - \frac{1}{4} x^2 = \sum_{k=1}^{\infty} (-1)^{n-1} \frac{\operatorname{Cos} nx}{n^2} \qquad (-\pi < x \leqslant \pi)$$

Faisant $nx = \pi$ et changeant les signes, il vient :

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = -\frac{1}{12}\pi^2 + \frac{1}{4}\pi^2 = \frac{\pi^2}{6}$$

2 - APPLICATION A

$$\varphi(t) = \Gamma(1+t)/(1+t)^{1+t} \qquad \varphi(o) = \Gamma(1)/1 = 1$$

$$\Phi(t) = \operatorname{Log} \varphi(t) = \operatorname{Log} \Gamma(1+t) - (1+t) \operatorname{Log}(1+t)$$

$$\frac{\varphi'(t)}{\varphi(t)} = \frac{d}{dt} \operatorname{Log} \Gamma(1+t) - 1 - \operatorname{Log}(1+t)$$

$t = 0$

$$\frac{\varphi'_o}{\varphi(t)} = -\gamma \qquad -1 \qquad \varphi'_o = -(1+\gamma) = \text{Moment d'ordre 1}$$

$$\frac{\varphi''(t)}{\varphi(t)} \quad \frac{\varphi'^2(t)}{\varphi^2(t)} = \frac{d^2}{dt^2} \operatorname{Log} \Gamma(1+t) - \frac{1}{1+t}$$

$t = 0$

$$\frac{\varphi''}{\varphi_o} - \frac{\varphi'^2}{\varphi_o^2} = \frac{\pi^2}{6} - 1 \iff \varphi''_o - \varphi'^2_o = \frac{\pi^2}{6} - 1 = \text{variance} = 0{,}645$$

3 - APPLICATION A

$$\varphi(t) = \Gamma\left(\frac{1}{2} + t\right) \Big/ \sqrt{2\pi}\left(\frac{1}{2} + t\right)^{\frac{1}{2}+t} \qquad : \varphi(o) = \sqrt{2}/\sqrt{2} = 1$$

$$\Phi(t) = \text{Log } \varphi(t) = \text{Log } \Gamma\left(\frac{1}{2} + t\right) - \left(\frac{1}{2} + t\right) \text{Log}\left(\frac{1}{2} + t\right) - \text{Log } \sqrt{2\pi}$$

$$\frac{\varphi'(t)}{\varphi(t)} = \frac{d}{dt} \text{Log}\Gamma\left(\frac{1}{2} + t\right) - \text{Log}\left(\frac{1}{2} + t\right) - 1$$

$$\frac{\varphi''(t)}{\varphi(t)} - \frac{\varphi'^2(t)}{\varphi^2(t)} = \frac{d^2}{dt^2} \text{Log}\left(\frac{1}{2} + t\right) - \frac{1}{\frac{1}{2} + t}$$

Pour $t = 0$, on a $z = -\frac{1}{2}$ dans $\Gamma\left(\frac{1}{2} + t\right) = \Gamma(1 + z)$; d'où

$$(1) \qquad \frac{\varphi_o'}{\varphi_o} = \left[-\gamma + \frac{-\frac{1}{2}}{1.\frac{1}{2}} + \frac{-\frac{1}{2}}{2.\frac{3}{2}} + \frac{-\frac{1}{2}}{3.\frac{5}{2}} + \ldots \right] + L\,2 - 1$$

$$= -\left[\gamma + 2\left(\frac{1}{1.2} + \frac{1}{3.4} + \frac{1}{5.6} + \ldots\right) \right] + L\,2 - 1$$

Démontrons que :

$$\frac{1}{1.2} + \frac{1}{3.4} + \frac{1}{5.6} + \ldots = \frac{1}{1} - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} \ldots = L\,2$$

En effet : $1 + x + x^2 + \ldots = (1 - x)^{-1}$ s'intègre et donne :

$$\frac{x}{1} + \frac{x^2}{2} + \frac{x^3}{3} + \ldots = -\text{Log}\,(1 - x)$$

Pour $x = -1$ la série est convergente et sa somme est $-L\,2$

$$\sigma^2 = \frac{\varphi_o''}{\varphi_o} - \frac{\varphi_o'^2}{\varphi_o^2} = \varphi_o'' - \varphi_o'^2 = \frac{d}{dt^2} \text{Log}\left(1 - \frac{1}{2}\right) - 1/1/2$$

$$= 4\left[\frac{1}{1^2} + \frac{1}{3^2} + \frac{1}{5^2} + \ldots\right] - 2$$

Faisant $x = 0$ dans (T) :

$$\frac{\pi^2}{12} - \frac{1}{4}x^2 = \Sigma(-1)^{n-1} \frac{\text{Cos } n\pi}{n^2}$$

Il vient :

$$\frac{\pi^2}{12} = \frac{1}{1^2} - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \ldots$$

Comparons à :

$$\frac{\pi^2}{6} = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \ldots$$

On en tire :

$$\frac{\pi^2}{8} = \frac{1}{1^2} + \frac{1}{3^2} + \ldots$$

d'où :

$$\sigma^2 = 4\left(\frac{\pi^2}{8}\right) - 2 = \frac{\pi^2}{2} - 2 = \frac{9,869}{2} = 2 = 2,935$$

4 - CUMULANTS D'ORDRES 3 ET 4 (Hypothèse F)

$$k_3 = \psi_o''' = \frac{d^3}{dt^3}\left[\operatorname{Log}\Gamma(1+t)\right]_o + (1+t)_o^{-2}$$

$$k_4 = \psi_o'''' = \frac{d^4}{dt^4}\left[\operatorname{Log}\Gamma(1+t)\right]_o - 2(1+t)_o^{-3}$$

$$\frac{d^3}{dt^3}\left[\operatorname{Log}\Gamma(1+t)\right]_o = \left[\frac{-2}{(t+1)^3} + \frac{-2}{(t+2)^3} + \ldots\right]_o = -2\left[\frac{1}{1^3} + \frac{1}{2^3} + \ldots\right] = -2A_3$$

$$\frac{d^4}{dt^4}\left[\operatorname{Log}\Gamma(1+t)\right]_o = +6\left[\frac{1}{1^4} + \frac{1}{2^4} + \ldots\right] = 6A_4$$

Le calcul direct donne :

$$A_4 \# 1 + \frac{1}{16} + \frac{1}{81} + \frac{1}{256} + \ldots + \frac{1}{10.000} = 1,068$$

$$A_3 \# 1 + \frac{1}{8} + \frac{1}{27} + \frac{1}{64} + \ldots + \frac{1}{1.000} + R = 1,1975 + R$$

avec $A_3 \# 1,20$

$$R \# \frac{1}{1331} + \frac{1}{1728} + \ldots + \frac{1}{10648} = 0,0036$$

d'où

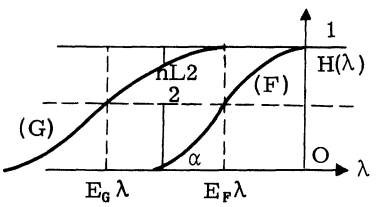
$$k_3 = -2,40 + 1 = -1,40$$

$$k_4 = +6,41 - 2 = 4,41$$

5 - CAS OU n SERAIT GRAND : RECHERCHE DE n OPTIMAL

Chaque distribution $H_F(\lambda)$, $H_G(\lambda)$ serait (asymptotiquement) de Laplace-Gauss.

La distance entre les deux $E(\lambda|F)$ et $E(\lambda|G)$ étant $\frac{n}{2}$ L2, on va la



représenter ($\forall n$) par un segment de longueur constante. Les écarts-types proportionnels à $\sqrt{n}$ ont alors des longueurs qui décroissent (l'un restant double de l'autre). La figure dépeint une situation raisonnable où λ sépare (F) de (G). Ceci suppose :

$$\frac{nL2}{2} > 2\sigma_F + 2\sigma_G \# 6\sigma_F$$

ou

$$6\sqrt{\frac{\pi^2}{6} - 1}\ \sqrt{n} < n\,\frac{L2}{2}; \quad \text{d'où} \quad \underline{n > 195}$$

6 - CAS OU n EST PETIT

On observe que (sous l'hypothèse F) les cumulants de $(x - Lx)$ sont :

$$k_2 = \frac{\pi^2}{6} - 1 = 0,645 \ ; \ k_3 = 1,40 \ ; \ k_4 = 4,41$$

donc :

$$k_3 \# 2k_2 \ ; \ k_4 = 3\,k_3 \# 6\,k_2$$

Pour une variable aléatoire "Gamma", on aurait :

$$\varphi_y(t) = E(e^{yt}) = \int_0^\infty e^{ty}\, e^{-y}\, y^{h-1}\, dy/\Gamma(h)$$

$$= (1 - t)^{-h}$$

$$\psi_y(t) = L\varphi_y(t) = -hL(1 - t) = ht + \frac{ht^2}{2} + \frac{2ht^3}{6} + \frac{6ht^4}{24} + \ldots$$

d'où $k_2 = h$, $k_3 = 2h$, $k_4 = 6h$ en toute rigueur.

On peut donc assimiler la distribution de $[x - Lx - (1 + \gamma)]$ à celle de $(y - h)$.

Par suite $2y$ ayant (comme on sait) la distribution de χ^2 avec $2h$ degrés de liberté, il s'en suit que $2(x - L_x)$ a une <u>distribution de χ^2 "non central" ou décentré</u>.

<u>Résultat</u> : La distribution de $x^2 - 4\lambda$ est un x^2 décentré.

En effet :

$$-4\lambda = 2\sum_{1}^{n}(x_j - Lx_j) - n2L(2\pi)$$

$$= \sum_{1}^{n} 2y_j + n[2(1+\gamma) - 2h - 2L\,2\pi] \quad \text{avec} \quad h = \frac{\pi^2}{6} - 1$$

$$-4\lambda - 2n\left[1 + \gamma - \frac{\pi^2}{6} + 1 - L2\,\pi\right] = \sum_{1}^{n} 2y_j = \sum_{1}^{n} \chi_j^2 = \chi^2$$

χ^2 ayant ainsi 2nh degrés de liberté.

Numériquement : $2h = 2 \times 0,645 = 1,3$; donc -4λ possède 1,3 n degrés de liberté :

$$2\left[1 + \gamma - \left(\frac{\pi^2}{6} - 1\right) - L2\,\pi\right] = -1,82 = \text{décentrage (à multiplier par n).}$$

Faisons par exemple n = 9 ; la Table du χ^2 avec 12 degrés de liberté donne :

$$\chi^2 = 18,549 \; ; \; 21,026 \; ; \; 24,054 \; ; \; 26,207$$

est significatif à :

$$\alpha = 10\,\% \;,\; 5\,\% \;,\; 2\,\% \;,\; 1\,\%$$

Avec le décentrage $-1,82\ n = -16,38$:

$$-4\lambda = 2,17 \; ; \; 4,65 \; ; \; 7,67 \; ; \; 9,84$$

est significatif à : 10 %, 5 %, 2 %, 1 %

A noter que l'échantillon $\mathcal{S}$ (étudié par M. Lerman) fournit $-4\lambda = 0,68$ qui correspond à $\alpha \neq 15\,\%$.

7 - RECHERCHE D'UNE LOI ASYMPTOTIQUE si $q_i \neq p_i$ $(n \longrightarrow \infty)$

En pratique, si N est assez grand et si les fonctions F, G sont continues, il est très naturel de supposer les q_i voisins des p_i ; d'où le calcul :

$$q_i = p_i(1+\varepsilon_i) \; ; \; L(q_i/p_i) = L(1+\varepsilon_i) = \varepsilon_i - \frac{\varepsilon_i^2}{2} + \frac{\varepsilon_i^3}{3} + \ldots$$

Sous (F), on a aussi : $n_i = np_i + n\eta_i$, d'où :

$$-\lambda = \Sigma n_i(L(q_i/p_i) = \Sigma np_i\varepsilon_i + n\eta_i\varepsilon_i - \frac{1}{2}\,np_i\,\varepsilon_i^2 + \ldots$$

Comme $\Sigma p_i = q$, $\Sigma q_i = 1 \Longrightarrow \Sigma p_i\,\varepsilon_i = 0$, il vient :

$$-2\,\frac{\lambda}{n} = -\,\Sigma 2\;\eta_i\,\varepsilon_i + \Sigma p_i\,\varepsilon_i^2 + \ldots$$

Si N est fixe et $n \longrightarrow \infty$, les écarts η_i sont de l'ordre de $\left(n^{-\frac{1}{2}}\right)$; et il reste :

$$= 2\,\frac{\lambda}{n} \;\#\; \Sigma p_i \varepsilon_i^2 = \sum \frac{(q_i - p_i)^2}{p_i}$$

ou

$$= 2\lambda \;\#\; \sum_i \frac{(nq_i - np_i)^2}{np_i}$$

Cette expression a une ressemblance formelle avec un χ^2 qui n'est pas l'effet du hasard comme on va le voir :

Si aucun des n_i n'est nul (ce qui suppose maintenant n supérieur à N, contrairement à ce qui avait lieu d'abord) les expressions suivantes sont définies :

$$\hat{I}_1 = \sum_1^N n_i \text{ Log } (n_i/n\, p_i) \;,\quad \hat{I}_2 = \sum_1^N n_i \text{ Log } (n_i/n\, q_i)$$

d'où :

$$2\lambda = 2\,\hat{I}_2 - 2\,\hat{I}_1 = 2 \sum_1^N n_i \text{ Log } (p_i/q_i)$$

Or $\hat{I}_1$ et $\hat{I}_2$ sont des informations de Kullback ; et on sait que :

sous (F)	$2\,\hat{I}_1 \sim \chi_1^2 = \sum \frac{(n_i - n\,p_i)^2}{n\,p_i}$	N − 1 degrés de liberté
sous (G)	$2\,\hat{I}_2 \sim \chi_2^2 = \sum \frac{(n_i - n\,q_i)^2}{n\,q_i}$	(idem)
sous (F)	$2\,\hat{I}_2 \sim \chi_c^2$, un χ^2 non central à N − 1 degrés de liberté	

Ainsi 2λ est la différence entre un χ^2 central et un χ^2 non central, sous (F), et aussi sous (G). Ceci ne nous permet d'ailleurs pas de trouver sa distribution tabulée (sauf cas particuliers dont nous parlerons dans un autre exposé)

$$2\lambda = \Sigma x_i^2 \left(\frac{1}{n\,p_i} - \frac{1}{n\,q_i}\right) \qquad (\text{avec } \Sigma x_i = n)$$

: 2λ est une combinaison linéaire de carrés de variables gaussiennes.

Remarque : En supposant tous les $\left(q_i - \frac{1}{N}\right)$ petits, on voit que $\mathcal{E}\lambda$ est sensiblement la moitié de $\mathcal{V}\lambda$, propriété des χ^2. La ressemblance ne semble pas s'étendre beaucoup plus loin (pour les fonctions caractéristiques).